Using the Method for Object-Based Diagnostic Evaluation (MODE) to Analyze Systematic Biases in the Global Forecast System (GFS)

T. Dorian[1,2], G. White[1], F. Yang[1,2], V. Tallapragada[1], T. Jensen[3] , and J.H. Gotway[3],
NOAA/NWS/NCEP/EMC[1], *I.M. Systems Group, Inc.* [2], *and NCAR/RAL Developmental Testbed Center*[3]

WHAT IS MODE?

MODE stands for "Method for Object-Based Diagnostic Evaluation" and is one of several forecast verification tools in the Model Evaluation Tools (MET) verification package developed by the National Center for Atmospheric Research (NCAR) Developmental Testbed Center (DTC). MODE was developed using precipitation as the forecast variable to be verified, but this object-based tool can also be applied to any variable that can be defined as an object on a gridded map.

Object-oriented verification techniques are beneficial to forecasters and model developers because these techniques provide diagnostic information on the differences between forecast and observations in terms of spatial displacement, coverage areas, orientation, and intensity. Object-based verification methods also better show the benefits of higher resolution models, as they avoid the "double penalty" problem that is so prevalent in traditional verification measures.

MODE works by comparing a gridded forecast file to a gridded observations file. The raw forecast and observation fields are put onto the same grid and are smoothed using a **user-defined convolution smoothing radius**, which is in units of grid squares. After the fields are smoothed, objects are defined wherever the values of the forecast variable of interest equal or exceed a **user-defined intensity threshold** (in variable units; millimeters in this case of precipitation). Once objects are identified based on the convolution radius and the intensity threshold and additional user-defined parameters in a configuration file, the original intensities within the raw forecast field are added back into the newly defined objects.

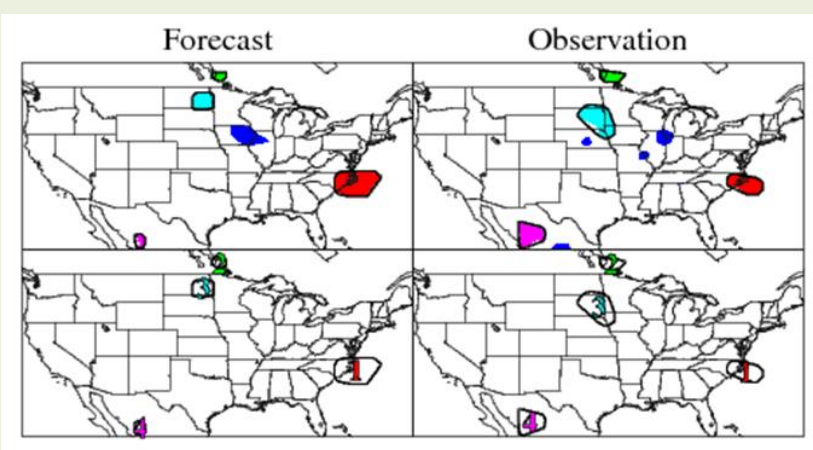

Figure 1: An illustration of the objects that MODE identified in the forecast (*left*) and observation (*right*) fields based on a user-defined convolution radius of 5 grid squares and an intensity threshold of >= 11 millimeters.

Following the identification of objects, MODE calculates simple object attributes such as centroid latitude/longitude, *n*th percentile intensity, object area, object axis angle, and object count. MODE also computes pairwise attributes such as centroid distance, percentile intensity ratio, area ratio, axis angle difference, union area, intersection area, and boundary distance for all possible pairs of objects.

Specific pairwise attributes are then considered in the computation of total interest values. The *interest value* is a summary measure that quantifies the overall similarity between two objects across fields and is computed using a fuzzy logic engine. User-defined weights for the pairwise attributes are included in the computation of the total interest values, T(α), as well as interest maps and confidence maps as shown in the equation below:

$$T(\alpha) = \frac{\sum_i w_i C_i(\alpha) I_i(\alpha_i)}{\sum_i w_i C_i(\alpha)}$$

Interest values range between 0 and 1, with values closer to 1 indicating a better match between forecast and observation objects. A forecast and observation object pair is considered to be a "match" if the interest value exceeds a user-defined interest threshold, which in this analysis is 0.7 (the default). MODE output includes statistics on simple objects, pairwise objects, and matched objects to evaluate how close the forecast and observations are to each other, which essentially examines how well the model performs.

ABOUT THIS STUDY

Verification used in this analysis include Climatologically Calibrated Precipitation Analysis (CCPA) files of 1/8° resolution for precipitation verification, and GFS and ECMWF analyses for the meridional jet verification.

The current operational GFS (T1534) horizontal resolution is 13km, or about 1/8°, however the forecast files used for precipitation verification in this study are 0.25° resolution. The current operational ECMWF files are similarly 0.25° resolution. For jet verification, the forecast and analysis files used were a coarser 1° resolution.

The precipitation forecasts and observations were put onto the NCEP grid 193, which is ¼° resolution over the contiguous United States. The meridional jet forecasts and analyses were put onto the NCEP grid 232, which is 1° resolution over the Northern Hemisphere.

In this study the periods analyzed were May 11-August 11 in summer 2016 for precipitation and a 3-month period (December, January, February) in winter 2016-2017 for jet verification.

All forecasts were initialized at 00Z. Precipitation forecasts and observations were grouped into 24-h precipitation accumulation periods. For precipitation verification, the focus was on forecasts for 24-h precipitation accumulations leading up to the 36, 60, 84, 108, 132, 156, and 180 hour forecasts. For meridional jet verification, the focus was similarly the 36, 60, 84, 108, 132, 156, and 180 hour forecasts.

RESULTS

The *Median of Maximum Interest* (MMI) is determined by finding the maximum total interest value associated with each individual forecast and observation simple object, and then finding the median within that set of maximum interest values.

MMI – median of maximum interest with respect to both forecast and observation objects
MMIO – median of maximum interest with respect to observation objects only (starts with obs)
MMIF – median of maximum interest with respect to forecast objects only

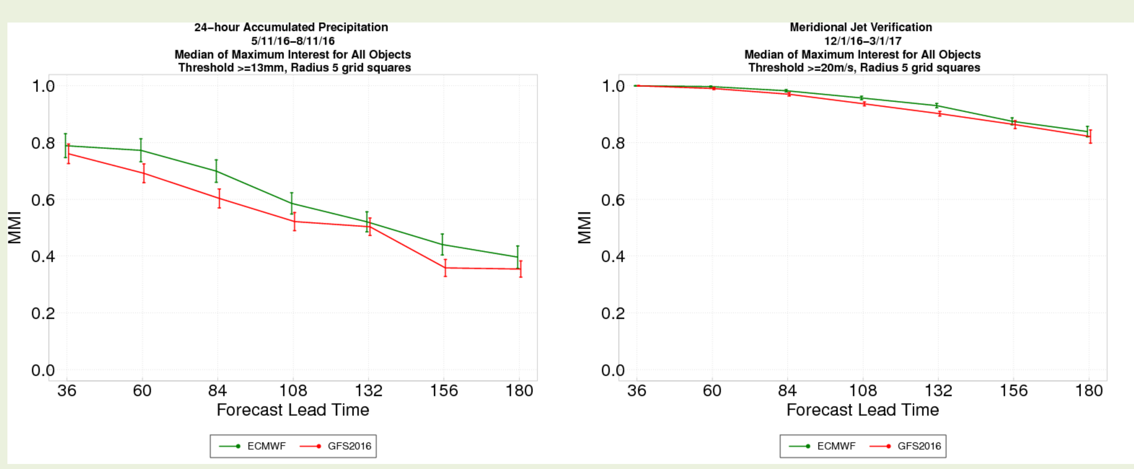

Figure 2: Median of Maximum Interest (MMI) for all precipitation (*left*) and meridional jet objects (*right*) as a function of forecast lead time. The GFS2016 (T1534) is in red and the ECMWF is in green.

The MMI values decreases with forecast lead time for both global models. The ECMWF has higher MMI values for both precipitation in the warm season and meridional jets in the cold season. Notice that the MMI values are much higher for jets than those for precipitation – this reflects the influence of object size in this object-based verification method and also reflects how the models better predict large-scale features.

RESULTS CONTINUED…

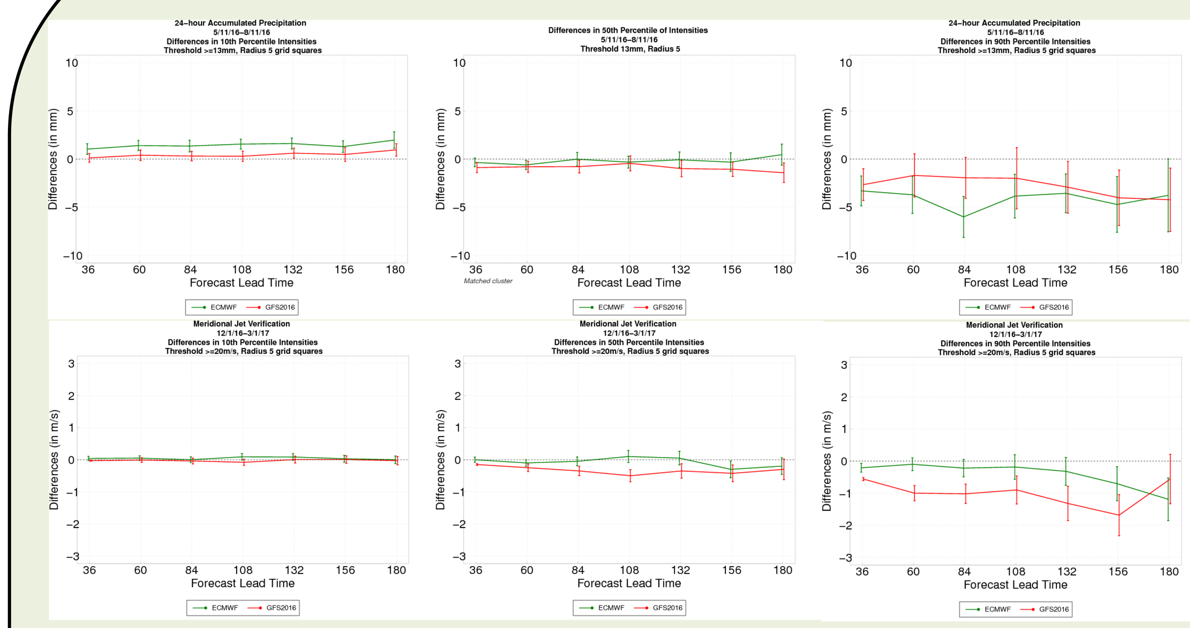

Figure 3: Differences in the 10th, 50th, and 90th percentile intensities for 24-hour accumulated precipitation between forecasts and CCPA observations during the May-August 2016 period (*top*) and for meridional jet between forecasts and analyses during the December-March 2016-2017 period (*bottom*). The ECMWF is depicted in green and the GFS2016 is depicted in red.

For precipitation accumulation forecasts, both the ECMWF and the GFS2016 show a positive intensity bias for the lightest precipitation, with the ECMWF having a larger positive bias. For median intensity precipitation, the ECMWF is closer to CCPA observations and the GFS2016 has a slight negative bias. For the heaviest precipitation, both models underestimate amounts, but the GFS2016 is closer to CCPA observations than the ECMWF. This implies that the GFS2016 did better for extreme precipitation cases.

For meridional jet intensity forecasts, both global models do fairly well with the weakest jets. For median intensity jets, the ECMWF is closer to jet analyses while the GFS2016 tends to be too weak. Lastly, for the strongest jets, both the GFS and the ECMWF underestimate the strength of the jets, but the GFS2016 has a much larger negative bias compared to the ECMWF.

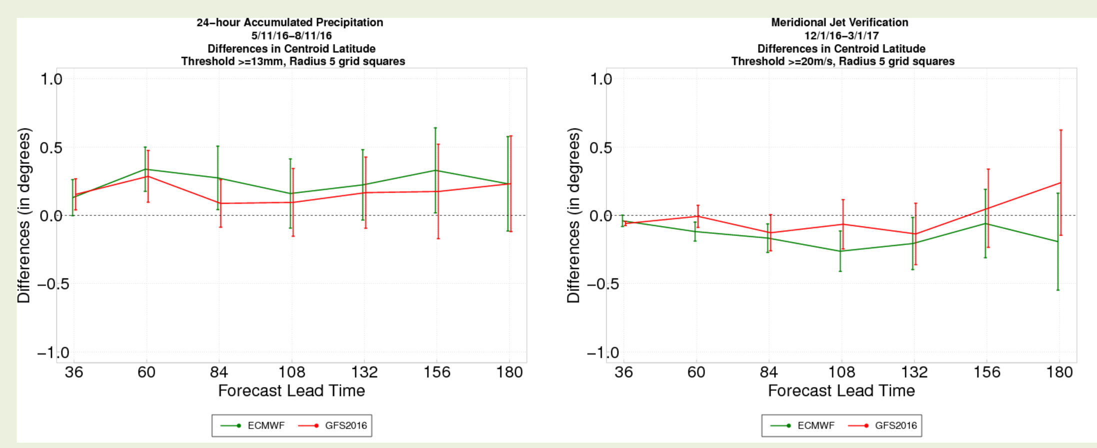

Figure 4: The differences in centroid latitude for 24-hour accumulated precipitation between forecasts and CCPA observations (*left)* and for meridional jets between forecasts and analyses (*right*). The GFS2016 is in red and the ECMWF is in green.

The GFS2016 and the ECMWF both have a northern bias for precipitation objects, and the differences between the two global models are not statistically-significant. For meridional jets, both the GFS2016 and the ECMWF have a southern bias, with the ECMWF having a larger southern bias, however differences are not statistically-significant.

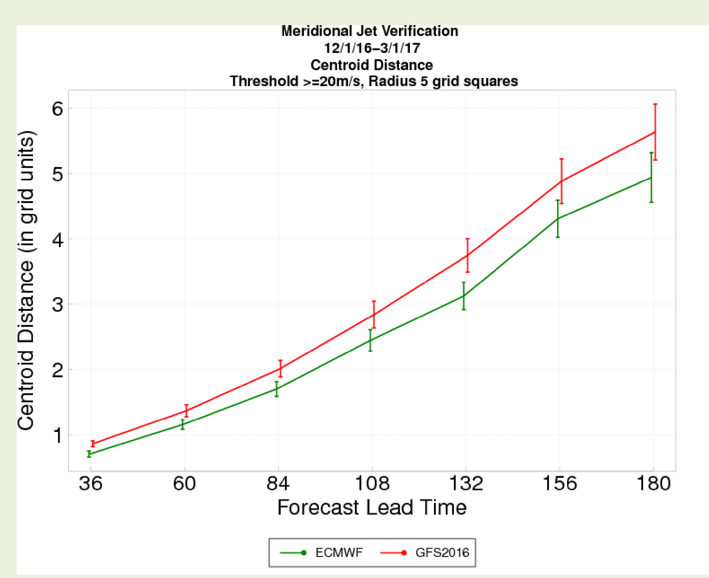

Figure 5: Centroid distances as a function of forecast lead time for meridional jet verification. Larger centroid distances mean that the forecast is further from verification. The GFS (in red) has larger centroid distances than the ECMWF (in green) for all forecast lead times, implying that the ECMWF had better position forecasts.

For more information on MODE and results, please visit my webpage at: **http://www.emc.ncep.noaa.gov/gc_wmb/tdorian/** or send an email to *tracey.dorian@noaa.gov*